

DOCUMENT RESUME

ED 265 679

EC 181 368

AUTHOR Hoge, Robert D.; Cudmore, Laurinda
TITLE The Use of Teacher-Judgment Measures in the
Identification of Gifted Pupils.
PUB DATE 85
NOTE 49p.; Support for this project provided by the
Pickering Fund.
PUB TYPE Guides - Non-Classroom Use (055)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Definitions; Elementary Secondary Education;
*Gifted; Research Methodology; Research Needs;
*Talent Identification; *Teacher Role

ABSTRACT

The paper intends to assess the psychometric properties of teacher-judgment measures for identifying gifted students in terms of the available empirical data. Problems of definition are cited, and the neglect and disparagement of teacher judgments as an alternative identification strategy are noted. The major conclusion from the review is that there is little basis for the negative assessments so often associated with these measures. Methodological flaws and weaknesses are cited. The paper includes a set of recommendations regarding the use of measures in the identification of gifted pupils (including expansion of the use of teacher judgments) and a set of recommendations regarding future research on the measures (including the need for more adequate definitions of the gifted construct). A nine-page bibliography concludes the paper. (CL)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

- ✓ This document has been reproduced as received from the person or organization originating it.
- !! Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

The Use of

1

ED265679

The Use of Teacher-Judgment Measures in the Identification
of Gifted Pupils

Robert D. Hoge and Laurinda Cudmore

Carleton University

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Robert D. Hoge

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

Running head: THE USE OF TEACHER-JUDGMENT MEASURES

EC181368

The Use of Teacher-Judgment Measures in the Identification
of Gifted Pupils

Policy concerning the treatment of gifted and talented children in schools has fluctuated widely over the years (Tannenbaum, 1979; Whitmore, 1980). The extremes of the continuum have been defined by a policy of equality, where no special treatment is provided the child with exceptionally high abilities, and a policy of special treatment for such pupils.

We are, at present, at the special treatment end of that continuum. This position is reflected in the United States in a federal law mandating the United States Department of Education to provide special attention to the needs of gifted and talented children (U.S. Pub. L. 91-230). There is also evidence that individual states are showing an increasing commitment to the expansion of gifted classes and other types of enrichment programs (Alvino, McDonnel, & Richert, 1981; Karnes & Collins, 1981; U.S. Department of Education, 1983). There is evidence for similar trends in the United Kingdom (Freeman, 1979) and in Canada (Borthwick, Dow, Levesque, & Banks, 1980). For example, the Province of Ontario has recently directed all provincial boards to initiate procedures for the identification and special treatment of gifted children (Ontario Legislature, Bill 82).

The existence of these special programs for the gifted create, of course, a need for identification procedures. If we are going to select out children for special treatment, we need some bases for making the selection decisions. The focus of this paper is on the use of teacher-judgment measures in this decision process. We begin, however, with a general discussion of problems of definition and identification associated with the selection of gifted children.

Problems of Definition

The goal in the educational setting is the identification of children with exceptional abilities who will profit from special educational programs. A first step in achieving this goal is the development of a definition of the "gifted" construct. That is, we must specify and define the qualities (traits and/or behaviours) which are associated with success in these programs.

We immediately encounter a problem in this respect, and the problem is that there is no real agreement here respecting the definition of such a construct. What we have are a large number of definitions which vary along several dimensions. Some good discussions of this definition problem are available (e.g., Fox, 1981; Getzels & Dillon, 1973; Renzulli, 1978; Rosenfield, 1983; Treffinger, Pyryt, Hawk, & Houseman, 1979; Tuttle & Becker, 1980), and we will touch on only two areas of variability especially relevant to our topic.

There is, first of all, variability among definitions with respect to the breadth of qualities or traits represented. At one extreme are those definitions which deal with the construct in terms of a single characteristic such as intellectual potential (e.g., Terman, 1925) or creativity (e.g., Torrance, 1965). At the other extreme are complex, multivariate definitions which include a broad range of traits or qualities. An example of the latter is the definition proposed by Hagen (1980). This definition includes 15 dimensions relating to cognitive characteristics (e.g., use of quantitative expressions and quantitative reasoning), academic skills (e.g., absorption in intellectual tasks), and personality characteristics (e.g., persistence on uncompleted tasks). There is evidence that the recent trend is toward multivariate rather than univariate definitions (Fox, 1981; Renzulli, 1978, 1984; Rosenfield, 1983), but there remains considerable confusion over the point in the literature.

A second dimension of variability relates to the nature of the qualities represented in the definitions. The focus has traditionally been on cognitive capacities, but there has often been controversy over the definition of these capacities and the relative weights to assign basic intellectual ability, academic achievement, and creativity. The scope of this controversy has been widened with recent efforts to include motivational and personality variables within the definitions (e.g, Renzulli, 1978, 1984). There is, in other words, little agreement or consistency in the literature respecting components of the giftedness construct.

Our main point is that there is no single definition of giftedness relevant to the school setting. What we have are a number of different definitions which vary widely as to scope and substance. It is important to keep this variability in mind, since it is relevant to our subsequent discussion of the evaluation of the adequacy of the teacher-judgment measures.

Alternative Identification Procedures

Instruments and procedures appropriate for the identification of gifted pupils should ideally be developed from a "gifted" construct. That is, having settled on the complex of traits and/or behaviors denoting gifted potential, we would then proceed to select or develop measuring instruments appropriate for assessing those traits or behaviors. This ideal procedure has generally not been followed. What has usually happened is that measures and identification procedures have been selected on the basis of availability or convenience, and a definition of the underlying construct, to the extent that this has been of concern at all, has followed from the measurement operations. This practice of allowing the choice of measuring instruments to precede the development of definitions has helped to contribute to the

definitional problem just discussed. It has also led to a proliferation of measures and procedures.

What we encounter is considerable variability with respect to the types of psychological measures employed in the identification of gifted children, the ways in which different measures are combined (if, in fact, more than one measure is used), and the procedures used for translating test scores into selection decisions. This situation has been discussed by a number of writers (Feldhusen, Asher, & Hoover, 1984; Fox, 1981; Hagen, 1980; Karnes & Collins, 1981; Rosenfield, 1983; Yarbrough & Johnson, 1983) and will not be dealt with in detail here. We will, however, outline the various kinds of measuring instruments employed in the identification of gifted children and make some comment on the extent of dependence on them.

The major categories include individual and group intelligence tests, individual and group achievement tests, tests of creativity, and teacher-judgment measures, including nomination and rating procedures. Other types of measures are sometimes encountered (e.g., peer ratings, personality inventories), but these constitute the major categories of identification instruments.

A number of researchers have reported data on the extent of dependence on these different types of measures in actual selection situations (Alvino et al., 1981; Borthwick et al., 1980; Jenkins, 1979; Yarbrough & Johnson, 1983). These surveys document that a variety of measuring instruments and decision strategies are used in the identification of gifted children, but they also make clear that the greatest dependence in these selection settings is on individual intelligence tests and on teacher-judgment measures. The use of intelligence tests in the selection of gifted children has been widely

discussed in the literature (Fox, 1981; Harrington, 1982; Sattler, 1982; Treffinger, 1984). The use of teacher judgments in this context has, however, been a somewhat neglected issue. It constitutes the major concern of this paper.

Attitudes Toward Teacher-Judgment Measures

The surveys cited above document the fact that there is a very heavy reliance on teacher judgments in the identification of gifted children. There is, however, something of a paradox here. The dependence on the judgments is accompanied by what appears to be a deep-rooted suspicion as to their worth. One frequently encounters this suspicion in discussions with school psychologists, psychometrists, educational researchers, and even, at times, teachers themselves.

There is also ample documentation for this negative evaluation of the judgments from within the literature. The following quotes are offered by way of illustration:

On occasion, teacher or peer nominations are accepted when no test scores are available, but this approach is considerably less valid ...Students selected by teachers tend to be those conforming to teacher guidelines and achieving well as a result. Only with proper guidelines or checklists do teachers begin to be even partially accurate in their selection procedures...

(George, 1979, p. 223)

Nomination by teachers is one of the most widely used and recommended means for identifying potentially gifted pupils, yet the method is of limited usefulness. Studies show that alone, teacher nomination proves the least effective screen.

(Borthwick et al., 1980, p. 18)

It is reasonable to hypothesize, then, that early identification of children with exceptional abilities is important. How, though, can we assure that such children are actually identified and helped? Several studies have reported that kindergarten teachers are woefully inaccurate in recognizing those children whose intellectual talents can be confirmed with intelligence tests.

(Robinson et al. 1979, p. 140)

While teacher nomination of gifted children is used more extensively than any other approach, it is successful only about 45% of the time in identifying gifted children.

(Sattler, 1982, p. 437)

These negative evaluations of the judgment measures are frequently supported by reference to one or more empirical studies which presumably demonstrate the inadequacies of the judgments. Gear's (1976) review article is also often cited in support of the negative assessments. Gear stated the following general conclusion from her review of five empirical studies relevant to teacher judgments of giftedness: "A review of the literature related to teachers' accuracy in the identification of gifted children indicates that teachers are relatively poor at this task" (Gear, 1976, p. 487).

The issue being raised here concerns, of course, the validity of the teacher-judgment measures. The popular assumption, as we have seen, is that the judgments are of very limited validity; that, in fact, they represent a poor basis for the selection of gifted pupils. The purpose of this paper is to treat that assumption critically and, to evaluate it in terms of the available empirical data.

It should be recognized that there are two senses in which this issue of

the validity of the teacher judgments is important. First, as we have seen, the judgments are widely used in making decisions about the placement of children into special classes or programs. These are important decisions so far as the child is concerned, and it is legitimate to raise a question about the quality of the decisions. Second, the negative assessments of the teacher judgments may, under some circumstances, be construed as criticisms of the competence of teachers. It is rare to find explicit criticisms of teachers in this literature. However, assessments of pupils constitute an important part of the teaching process. If it can be shown that teachers are poor in this respect, then there is, in fact, a basis for criticism.

Validity Data

Our assessment of the teacher-judgment measures is based on a review of empirical studies in which data have been presented on the psychometric properties of this type of measure. Table 1 presents a summary of the

Insert Table 1 about here

results of these studies along with information respecting the type of sample employed, the type of predictor and criterion measure used, and the nature of the analysis.

Characteristics of the Studies

The judgment measure: Two types of judgment measure are represented in the studies reviewed. First, there are the nomination procedures where the teacher is asked to identify pupils satisfying a particular definition (e.g., "intellectually gifted"), and, second, there are the rating procedures where the teacher is asked to rate the pupils with respect to one or more dimensions relevant to giftedness.

A close examination of these studies reveals that there is, in fact, considerable variability within these two categories. So, for example, the studies employing nomination procedures show variability in the way in which the nomination category is defined and, more important, in the precision with which it is defined for the teacher. Thus, Gear (1978) provided teachers (in one of her groups) with some training in the identification of gifted potential before asking for the nominations. Most of the studies, on the other hand, have simply involved asking teachers to identify their "intellectually gifted" or "mentally gifted" pupils without providing any guidance in what is meant by the categorization.

There is also variability among the studies employing rating measures since a variety of rating dimensions and formats are represented in the research. So, for example, Ashman and Vukelich (1983) employed a 26-item scale tapping various cognitive and academic areas of competence. Scores from the scale were then combined into a single composite score for purposes of analysis. Borland (1979) has developed an interesting checklist measure which contains 15-items tapping a variety of aspects of gifted potential (e.g., "reasons things out independently", "reads a great deal, usually well beyond grade level"). A scale developed by Renzulli and Hartman (1971) deserves special note because it is being used increasingly in the identification of gifted pupils. The Scale for Rating Behavioral Characteristics of Superior Students (SRBCSS) is a 37-item rating scale yielding scores relevant to learning, motivational, creativity, and leadership characteristics. These examples illustrate broad-range types of measures which tap a variety of dimensions thought relevant to giftedness. Other measures here provide for a focus on a single dimension of giftedness such as creativity or achievement.

Criterion measures: A variety of types of criteria are represented in the studies summarized in Table 1. The most frequently employed type is the individual intelligence test, with 13 of the 22 studies using such a test as the sole criterion measure or as one in a battery of criteria. The Stanford-Binet and the WISC-R are the most commonly used tests. Nine of the studies included a creativity test as a criterion measure. This generally involved the use of a standardized type of measure (e.g., the Wallach-Kogan Tests of Creativity), but self and peer ratings of creativity are also represented. Three of the studies included a standardized achievement test as a criterion index, while in one case expert judgments constituted the criterion measure.

Some comment on the adequacy of these criterion measures is perhaps appropriate at this point. The intelligence achievement tests employed in these studies are standardized instruments with known, and generally sound, psychometric properties. The creativity measures are, on the other hand, more experimental instruments and some questions exist with respect to their psychometric properties (cf. Anastasi, 1983; Sattler, 1982). There is also an issue to be raised with respect to the appropriateness or relevance of these criterion measures in the assessment of gifted potential. That issue will, however, be raised later in the paper.

Design and analysis: All of the studies included in Table 1 employed a correlational design, although the type of statistical analysis used within the design varied somewhat. Most of the researchers established relations between judgmental measures and concurrently collected criterion measures by means of correlational statistics. The most significant departures so far as the timing of data collection is concerned occur with the Harrington, Block, and Block (1983) and Klausmeier, Harris, and Ethnathios (1962) studies in which criterion information was actually collected prior to collection of the

judgment scores. A second form of data analysis is seen in the Chambers, Baron, and Sprecher (1980) and Dewing (1970) studies. They employed what is basically a correlational design, but they formed quasi-experimental groups on the basis of teacher judgments and compared those groups by means of t-tests in one case and Chi-Square tests in the other.

There is yet a third type of analysis encountered in this set of studies. As can be seen from the table, nine of the researchers employed indices of prediction efficiency and/or effectiveness in their analyses. These provide estimates of the degree of accuracy of the teacher judgments relative to a criterion. The values for these indices are calculated according to the following formulae:

$$\text{Effectiveness} = \frac{\text{number of confirmed gifted identified}}{\text{number of confirmed gifted}}$$

$$\text{Efficiency} = \frac{\text{number of confirmed gifted identified}}{\text{total number identified}}$$

The effectiveness index reflects the ratio of the number of pupils nominated by the teacher as gifted relative to the total number identified as gifted on the basis of the criterion measure. The efficiency index reflects the ratio of successful teacher designations relative to the total number identified by the teacher. In terms of the decision accuracy model, the effectiveness index reflects the ratio of true positives to the total of true positives and false negatives, while the efficiency index reflects the ratio of true positives to the total of true and false positives. This represents, of course, a legitimate approach to the establishment of validity. We will see later, however, that there are some serious problems with the way in which this analytic procedure was applied in the present set of studies.

Results

The results of these studies are summarized in the last column of Table 1. An effort was made in producing this summary to present the most basic analyses respecting the judgment-criterion relations. In some cases this involved collapsing across groups of subjects or otherwise combining results. It should also be noted that the emphasis in this section of the paper is on a description of the judgment-criterion relations. The implications of these results for the validity of the teacher-judgment measures is dealt with in a later section.

Nomination studies: Most of the studies using a nomination type of judgment measure employed an efficiency-effectiveness index in their analysis. Further, with the exception of the Gear (1978) study which employed expert judgments, all of these studies used an individual intelligence test as the criterion measure. The effectiveness indices reported in this set of studies ranged from 0% to 86% with a mean value of 40%. This means that, on average, 40% of the children meeting the criterion of giftedness (as determined by the intelligence test score) were identified as gifted by the teacher. The reported efficiency indices ranged from 4% to 78% with a mean of 36%. This means that, on average, 36% of the children identified as gifted by the teacher actually met the criterion of giftedness.

Taken on one level, these results may be interpreted as reflecting moderate and highly variable levels of accuracy for the teacher judgments of gifted potential. However, there are some problems with the way in which this accuracy analysis has been applied in the present case, and the problems are of sufficient seriousness to lead us to question whether these results have much utility at all in the assessment of these judgment-criterion relations.

These efficiency-effectiveness indices are based on the proportions of correct decisions as derived from the judgment measure. The proportions are, however, affected by base rates and by chance occurrences, two factors which have received inadequate treatment in the studies.

Base rate refers to the proportion of subjects in the sample who meet the criterion of success. In the present case the reference is to the proportion of subjects who fall above the intelligence score cutoff. Base rate is determined by the nature of the sample and by the level at which the criterion cutoff is set. There is a direct link between decision accuracy and base rate since decision accuracy approaches a maximum as the base rate approaches 50%.

There are three points to be made with respect to the treatment of the base rate variable in this set of studies. First, most of the researchers are deficient in reporting base rates. Second, the criterion cutoff points are determined in a purely arbitrary manner and vary from study to study. The cutoff is whatever intelligence test score the researcher selects as the cutoff. Third, there is reason to believe that, in many of these studies at any rate, we are dealing with extreme base rates; that is, samples in which there are exceptionally high or exceptionally low numbers of subjects meeting the criteria. To the extent that we are dealing with unknown but probably extreme base rates, the accuracy indices which are being reported must be interpreted with great caution.

There is also a failure here to consider the operation of chance occurrences within the decision matrices. The efficiency and effectiveness indices provide us with estimates of decision accuracy relative to a criterion. They do not, however, tell us anything about the statistical

significance of particular accuracy levels in particular situations. That information must be provided through statistical tests (Meehl & Rosen, 1955), and such tests have not been reported in this set of studies.

These issues have been pursued at some length because we are dealing here with an important set of studies; these are the studies generally cited as supporting the negative evaluations of teacher judgments of giftedness. We will raise some other objections to these studies later in the paper when we talk about the implications of the results for validity, but we have tried to show here that the investigations are seriously deficient in terms of analytic procedures, and that the results must be used with great caution as sources of information about judgment-criterion relations.

There are two other results from the nomination studies which bear mention. Ashman and Vukelich (1983) reported correlations between nomination scores and intelligence test scores in addition to accuracy indices. A correlation of $r = .36$ obtained between the two measures. Dewing (1970) collected giftedness nominations from teachers, and used standardized creativity tests and peer ratings of creativity as criteria. Chi-square tests indicated significant relations between the nomination variable and the criterion measures.

Rating studies: Twelve of the studies summarized in Table 1 included a judgmental measure based on a rating format. Three types of criterion measure were represented in the studies: intelligence test scores, achievement test scores, and creativity test scores.

Six of the studies summarized there involved the use of intelligence test scores as criterion measures. Several rating formats were used in those studies, but it is interesting to note that, with one exception,

statistically significant relations were reported between the teacher rating measure and IQ scores. Two of the cases reporting significant relations employed the SRBCSS as a predictor measure. Thus, Ashman and Vukelich (1983) reported significant correlations between a composite SRBCSS score and intelligence test scores, while Renzulli, Hartman, and Callahan (1971) reported significant correlations between the standardized test scores and the Learning and Motivation subscales of the SRBCSS. Similarly positive results were also reported for alternative kinds of rating scales by Borland (1979), Chambers et al. (1980), and Kirk (1966). Ashman and Vukelich (1983) included a second rating measure in their study, and it, too, displayed a significant relation with the IQ criterion. The only negative result here was reported by Rust and Lose (1980) who failed to establish significant relations between SRBCSS subscale scores and intelligence test scores. It is worth noting, however, that a very restricted range of predictor and criterion scores were represented in their sample, and this may have contributed to the negative results.

Two of the studies included standardized achievement test scores as criteria, and in both cases statistically significant relations were reported between the judgment and criterion measures (Renzulli et al., 1971; Swenson, 1978). By way of illustration, Renzulli et al. (1971) reported correlations ranging from $r = .41$ to $r = .57$ between the Learning subscale of the SRBCSS and achievement score and r_s ranging from .42 to .60 between the Motivation subscale and the achievement scores.

The situation is somewhat more confused with those studies reporting relations between rating measures and creativity measures. Nevertheless, the results are instructive. Basically negative results were reported between

judgment and creativity measures by Mayfield (1979) and Swenson (1978). Mixed results were reported by Chambers et al. (1980) and Renzulli et al. (1971). The former researchers found significant results for some grade levels and some creativity test subscales but not for other grades or subtests. Renzulli et al. (1971) reported correlations between creativity ratings from the SRBCSS and creativity test subscale scores ranging from $r = .24$ to $r = .48$. Three of the seven correlations were statistically significant for the small sample of subjects involved.

There are, however, several studies reporting strong relations between judgmental measures and creativity test scores. Thus, Cunningham et al. (1979) reported significant relations between teacher ratings of pupil creativity levels and creativity test scores and similarly significant relations between judgments of the extent to which a pupil belonged in a gifted class and the test scores. Klausmeier et al. (1962) reported correlations between teacher ratings of expressional fluency, ideational fluency, and originality and corresponding scores from standardized tests. The correlations were, on the whole, statistically significant. The design of the Harrington et al. (1983) study differed somewhat from the others in that the creativity test scores were collected six years before the teacher ratings of creativity. Nevertheless, a statistically significant correlation was reported between the two indices of creativity. One final result to be mentioned here derives from the Davis and Rimm (1977) study where significant correlations were reported between teacher ratings of creative potential and a self-report measure of creativity.

There are some other analyses reported in connection with these rating measures which bear mention. Reliability data here are rather sparse, but

Renzulli et al. (1971) have reported test-retest and inter-rater agreement coefficients for the subscales of the SRBCSS. The stability coefficients range from .77 to .91, and the inter-rater agreement coefficients from .67 to .91. Stability coefficients have also been reported by Borland (1979) in connection with the rating scale which he developed. This analysis was based on ratings provided by two sets of teachers and separated by a two year interval. The reported coefficient was .86.

Some data relevant to the internal structure of the SRBCSS have also been reported. Burke, Haworth, and Ware (1982) presented a factor analysis of data collected with the four subscales of the rating instrument. Their analysis failed to support the validity of the four-factor structure claimed for that instrument; a single factor was shown to account for a significant amount of the variance. One limitation associated with that study should, however, be noted. These researchers were dealing with a group of pupils preselected as highly gifted, and their analysis was based, therefore, on a very restricted range of scores. A more positive result for the SRBCSS has been reported by Ashman and Vukelich (1983) who showed a significant relation between the total score of the SRBCSS and a composite giftedness score based on an alternative rating measure.

The result of item analyses reported in the recent Harrington et al. (1983) study are also of interest. Those researchers collected teacher judgments by means of the California Child Q-Set, a well standardized judgmental measure of personality and cognitive attributes. As noted above, significant relations were reported between the creativity subscales of that measure and standardized creativity tests. The authors also explored relations among the various subscales of that instrument in an effort to

assess the extent to which a judgmental construct of creativity could be said to exist independent of a general factor. Their results indicated that, in fact, the creativity ratings were associated with dimensions logically related to creativity and not significantly associated with divergent traits.

Implications for Validity

Construct validity: We will define construct validity as the extent to which a measure represents a meaningful and accurate index of an underlying construct (Cronbach & Meehl, 1955; Messick, 1981). In the present case we are asking to what extent the various nomination and rating procedures represent meaningful indices of gifted potential, the construct of primary interest here. We encounter a serious problem in this respect, for, as we have seen, there is no formal and explicit giftedness construct presented in this literature. Further, there is such a wide variety of operational definitions represented in the measurement procedures used in this research, and the operationalizations are frequently so unclear, that there is little basis for deriving a construct inductively from the research literature.

Nevertheless, there are some data which may be discussed in connection with this issue of construct validity. Two researchers have presented data respecting inter-correlations among components of rating measures. We saw that Burke et al.'s (1982) factor analysis of scores from the SRBCSS produced inconclusive results so far as the identification of meaningful factors within that measure were concerned. On the other hand, Harrington et al.'s (1982) internal analysis of the California Q-Set measure provided evidence that a meaningful judgmental construct of creativity existed within that measure.

A second procedure for assessing construct validity involves exploring

relations between components of a measure and scores from parallel measures with which those components are logically related. Some of the analyses reviewed in the previous section clearly fall within this category and provide support for the construct validity of the measures in question. For example, Ashman and Vukelich's (1983) demonstration of significant relations between composite measures from two teacher rating scales of gifted potential may be said to provide evidence that a meaningful giftedness construct exists, although it must be acknowledged that the alternative measures were being collected from the same group of teachers. Also relevant here are those demonstrations of significant relations between creativity ratings and creativity test scores (Cunningham et al., 1978; Harrington et al., 1983; Klausmeier et al., 1962; Renzulli et al., 1971). The Harrington et al. study is particularly interesting because they were able to provide some evidence of the convergent and discriminant validity of the creativity judgments, although formal tests of those forms of validity were not provided.

There is another set of studies which are often cited as relevant to the construct validity of the judgments, although their actual relevance to that issue is in doubt. The reference is to those studies involving the establishment of relations between teacher nominations and intelligence test scores. Strictly speaking, these analyses are relevant to criterion-related validity. However, they are often interpreted as having a bearing on construct validity. Thus, all of the negative assessments of teacher judgments which were quoted at the beginning of this paper involved reference to one or more of these nomination studies. Further, Gear's (1976) widely quoted conclusion respecting the inaccuracy of the judgments is based solely on these studies.

We saw in an earlier section that there are serious methodological flaws associated with these nomination studies. An additional point to be made here is that there is little basis for interpreting the results of these studies as having any bearing at all on construct validity. Two arguments will be presented in support of this view. First, it will be asserted that scores on an intelligence test represent too narrow a criterion against which to evaluate the accuracy or meaningfulness of a global index of gifted potential. There remains, as we have seen, great uncertainty over the definition of this construct of gifted potential, but there seems rather general agreement that the construct involves something more than intellectual or cognitive competence as assessed by an IQ test. Second, there is a problem here with respect to the degree of correspondence between predictor and criterion measure. The assessment of construct validity through the examination of relations between parallel measures depends on the assumption that the measures are, in fact, parallel. There is a failure to satisfy this condition in this set of studies. What we have here is a pairing of a global, vaguely defined judgmental measure with a very specific criterion measure. What is happening is that the researcher is inviting the teacher to formulate his or her own definition of giftedness as a basis for the nominations, but is then evaluating the judgment against the specific criterion of intelligence test scores. This procedure not only renders the results of questionable relevance so far as assessing construct validity is concerned, but it also appears somewhat unfair to teachers.

Criterion-related validity: Most of the studies reviewed in Table 1 which are not directly related to construct validity may be considered relevant to concurrent validity. That is, they involve efforts to relate a

judgmental measure to an alternative type of measure, with the two sets of measures collected at approximately the same time.

All of the nomination studies belong in this category, although, as we have seen, there have been efforts to interpret them as relevant to construct validity. In any case, most of the studies employing a nomination measure have shown only weak relations between judgmental and criterion measures. We have seen, however, that these studies are flawed in a number of respects, and, hence, they are of limited value so far as the assessment of criterion-related validity is concerned.

The studies employing rating procedures are generally more sound from a methodological point of view, and it is interesting to observe that they present a more positive picture so far as the concurrent validity of the judgmental measures is concerned. Thus, generally significant relations were reported between various rating formats and intelligence test scores (Ashman & Vukelich, 1983; Borland, 1979; Chambers et al., 1980; Kirk, 1966; Renzulli et al., 1971a; Swenson, 1978) and standardized achievement test scores (Renzulli et al., 1971; Swenson, 1978). The correlations are not always of high magnitude and there was a failure to establish validity (Rust & Lose, 1980), but, on the whole, statistically significant correlations have been established here.

These results relating to the concurrent validity of the judgmental measures are of some interest to the extent that they contribute to our understanding of the judgmental construct underlying the measures (Cook & Campbell, 1979; Cronbach & Meehl, 1955; Messick, 1981). For example, a finding that teacher ratings of mental giftedness relate significantly to scores from standardized achievement tests provides us with some information about the nature of that particular kind of rating measure.

It must also be recognized, however, that this type of concurrent validity information is of very limited utility from the point of view of making use of these measures within the educational setting. The teacher judgment measures of giftedness and creativity are used as sources of information in deciding whether or not to place pupils in special programs or classes for the gifted. It follows that the primary basis for evaluating the measures should be in terms of their predictive validity. In other words, we should be assessing the extent to which the measures are effective in identifying children who ultimately succeed or fail within these classes and programs. In answering this question, there is no substitute for predictive validity studies, and it is important to observe that not a single predictive validity study has been reported in connection with these judgmental measures.

Conclusions from the Review

The objective of this review was an evaluation of the psychometric properties of these teacher-judgment measures of gifted potential. A more specific concern was with the soundness of the negative evaluation so often associated with this type of selection tool.

The major conclusion to be drawn from the review is that the psychometric qualities of these judgmental measures have been neither extensively nor adequately tested. It may be noted, first, that we are dealing here with a relatively small sample of studies showing considerable variability with respect to the definition and operationalization of variables, design, modes of analysis, etc. Second, as we have seen, many of these studies are flawed in terms of design and analytic procedures. The consequence of these points is that there is, in fact, little basis here for

any conclusive statements about the reliability or validity of these measures.

It follows too that there is very little empirical foundation for the negative evaluations so often associated with these measures. There may be some grounds for such an evaluation from the clinical experience of educational workers (and such evidence deserves consideration), but there is certainly no firm foundation within this empirical literature. In fact, the thrust of more recent research seems to be in a direction supportive of the judgments (e.g., Ashman & Vukelich, 1983; Borland, 1979; Cunningham et al., 1978; Harrington et al., 1983), but here too the empirical base is limited.

Recommendations

Practical Considerations

One conclusion which might follow from an examination of this literature is that we should simply suspend all efforts at identifying gifted children until we are able to develop some improved assessment tools. Such a strategy is, of course, unavailable to us. Decisions are being made about the placement of children into these programs, and we must confront that reality. Further, in spite of the somewhat uncertain state of our current assessment tools and our knowledge of them, it seems to us that there are some clear lessons here for psychologists and others involved in making these decisions.

Our first recommendation is that those involved in the selection of gifted children should attempt to deal more adequately with the question of definition than has sometimes been the case in the past. Explicit guidelines must be developed and stated with respect to the traits, behaviors, and/or aptitudes which constitute the targets of the selection process in particular situations. The common practices of deliberately leaving the definitions

open or of allowing definitions to be dictated by the measuring instruments employed are not satisfactory. They lead to bad decisions and to unnecessary conflicts with parents. It must be acknowledged that there is little concrete guidance to be offered by the current empirical literature in the development of such definitions. There is, however, a good deal of valuable knowledge and advice represented in both the theoretical and empirical literatures, and practitioners will benefit from a familiarity with that information. The Hagen (1980) and Tuttle and Becker (1980) books are particularly rich sources of ideas respecting the definition of giftedness.

Our second recommendation is that the use of teacher judgments in the identification of gifted children should be continued, and, in fact, expanded. This may appear to be paradoxical advice given the conclusions of the previous review. There are, however, several considerations which lead us to this recommendation.

Our first argument is based on theoretical considerations. We have in the case of the classroom teacher a trained professional who has had extensive and varied interactions with the child. The teacher represents, potentially at least, an extremely valuable source of information regarding the qualities of the child. Our second argument is that, while there are undoubtedly limitations associated with these judgments, there are limitations associated with all of the types of measures used in the selection of gifted children, including intelligence tests (cf. Fox, 1981; Harrington, 1982) and creativity tests (cf. Barron & Harrington, 1981; Getzels & Dillon, 1973). Our third argument in favor of the use of these teacher-judgment measures is based on growing empirical evidence that, under optimal circumstances at any rate, teachers are capable of providing accurate

information about characteristics of pupils. This evidence comes from within the gifted literature (e.g., Borland, 1979) and from other judgmental areas (e.g., Hoge & Butcher, 1984; Rubin & Clark, 1983).

Our recommendation that we continue to depend on teacher judgments in this selection process is based, however, on three conditions. The first of these is to the effect that teachers should be given adequate preparation before providing the judgments. This means they should be fully familiarized with the purposes of the identification process. It also means that where nominations are being asked for the teacher is provided with an explicit definition of the "gifted" construct, or, where ratings are asked for, the teacher is provided some background in the use of the rating instrument. The importance of training in the collection of judgmental information has been emphasized by a number of writers within the gifted literature (Gear, 1978; Pledge, 1982; Schlichter, 1981) and in more general terms in the personnel psychology literature (see, for example, Borman, 1979; McIntyre, Smith, & Hassett, 1984).

The second condition to be met here is that teachers must be provided adequate tools for expressing the judgments. This practice of depending on ill defined nomination categories or ad hoc rating scales is not satisfactory. What is needed, in our opinion, is the development and use of standardized rating measures with known psychometric properties. We have seen that some efforts in the development and use of such instruments are being made. For example, the SRBCSS (Renzulli et al., 1971) represents a multivariate rating instrument developed from some theoretical considerations respecting gifted potential. Further, there have been efforts at assessing the psychometric properties of the instrument (e.g., Burke et al., 1982;

Renzulli et al., 1971) and at the development of normative data (Argulewicz, Elliott, & Hall, 1982). Other efforts at developing teacher rating measures of gifted potential have been reported by Ashman and Vukelich (1983), Borland (1979), Dirks and Quarforth (1981), Pledge (1981), and Rubenzer (1979), and practitioners are advised to familiarize themselves with those efforts.

The third condition we will state is to the effect that teacher judgments should be used in combination with other assessment tools in this selection process. We believe that, with proper preparation and effective tools, teachers can provide useful information with respect to the potentialities of children. Still, there will always be some limitations associated with the information (as there will be limits associated with all types of measures), and, therefore, the judgments should represent one of a number of sources of information within the selection situation.

Our third recommendation follows from this last point. This recommendation is to the effect that those professionals involved in this selection process should seek more adequate decision models for combining information from multiple sources and for translating scores from instruments into actual placement decisions. Many of the practices followed are simply too arbitrary and too simplistic, a point noted by many writers, including Feldhusen et al., 1984, Rosenfield, 1983, Treffinger et al., 1979, and Tuttle and Becker, 1980. Unfortunately, there are no well tested procedures available, but useful beginnings have been made by Feldhusen, Baska, and Womble (1981), Renzulli, Reis, and Smith (1981), Tuttle and Becker (1980), and within the Talent Search Project (Fox, 1981; Stanley, 1976; Stanley, Keating, & Fox, 1974).

Directions for Research

A familiarity with the research literature on the assessment of gifted

pupils can perhaps lead to despair over the fragmentary and sometimes methodologically weak approaches which are represented there. Our view is, however, more positive. We feel that there are few areas of applied psychology where there are so many opportunities for theoretically meaningful research which may also have immediate practical impact within the schools. We won't attempt a complete review of these research areas, but will indicate some research directions relevant to our topic of teacher judgments.

First, it is clear that the educational researcher has a contribution to make in the development of more adequate definitions of the "gifted" construct. In fact, a lack of empirical input has been a serious problem in the development of a definition. There are a number of directions from which this information will come. It may be expected, for example, that research on the construct validity of existing instruments will yield relevant information respecting the components of gifted potential. This type of research will be discussed below. A second direction might involve the systematic collection of perceptual and attitudinal data from teachers and other professionals with extensive contact with gifted children. This approach has been used with some success in the development of definitions of social and academic competence (e.g., Murphy, Jenkins-Friedman, & Tollefson, 1984; Kornblau, 1982). A third possible direction for this research stems from work on the analysis of teacher decision-making and judgmental processes (e.g., Borko & Cadwell, 1982; Shavelson, Cadwell, & Izu, 1977). All of these represent potentially useful approaches for the generation of information about the components of gifted potential.¹

A second area in which there is clear need for more research concerns the development of improved judgmental tools and improved educational programs in the use of those tools. This research will undoubtedly involve the development of new rating instruments, but it should also involve investigations of alternative rating formats, variations in the wording of scale items, etc. A few efforts to deal with these technical kinds of issues have appeared in this gifted literature (Ashman & Vukelich, 1983; Kirk, 1966), but the efforts have been rather sparse. There are, however, signs of increased research activity respecting rating technology within the personnel psychology literature (e.g., Borman, 1979; Imada, 1982; Love, 1981), and we should take advantage of those models. The effects of training on the quality of the gifted judgments has also been a neglected issue even though Gear (1978) demonstrated a number of years ago that a training program can enhance decision quality. Here again there are some useful research models in the personnel psychology literature where there is considerable interest in the effects of alternative training programs on the quality of judgmental assessments (e.g., McIntyre et al., 1984; Zedek & Cascio, 1982).

Our third recommendation is to the effect that increased research attention should be directed toward the measurement properties of teacher judgment measures. It is a little curious that we concern ourselves so much with the reliability and validity of standardized tests and observational measures and yet adopt such a casual attitude when it comes to judgmental measures (Hoge, 1983, 1984). There is, however, no justification for that casual approach, especially where the measures are being used as identification or selection tools. The Standards for Educational and Psychological Tests (American Psychological Association, 1974) is quite clear

on the point: "It is intended that these standards apply to any assessment procedure, assessment device, or assessment aid; that is, to any systematic basis for making inferences about characteristics of people" (p. 2).

There are a number of specific areas in which research is needed on the psychometric properties of these measures. First, as we have seen, relatively little attention has been paid to the issue of reliability. The effective use of the measures in applied and research settings depends on precise information about their consistency and stability, and more information of that sort must be reported.

Second, there is a clear need for increased attention to the construct validity of the existing measures. Information about the meaning of the measures seems particularly important in this case where the instruments are being used not only as a basis for decisions about the child, but where a labeling process is also implicitly involved. This construct validity research might take two directions. The first approach would involve factor analyses of data collected with the various instruments. We have seen from the child pathology literature that this can be a very fruitful approach for analyzing the meaning of scores from instruments and for refining constructs (cf. Achenbach & Edelbrock, 1978; Edelbrock, 1979; Quay, 1979). The second approach would involve efforts to relate scores from the judgmental measures to parallel scores from alternative instruments. Twenty years ago Adams (1964) presented a model of a multitrait-multimethod investigation of a giftedness measure. That would still be a very useful study.²

The collection of information relevant to the concurrent validity of the judgmental measures should continue. We have seen that such information is

of limited utility so far as evaluating the use of these instruments as selection devices is concerned, but we may expect that the information will ultimately contribute to a better understanding of the measures and of the giftedness construct.

The critical need here is for predictive validity studies, and this constitutes our major recommendation. We have seen that scores from these judgmental measures are used in making decisions about the placement of children into special classes or special programs. There is, then, an implicit assumption that scores from the measures are predictive of success or failure within these programs. The absence of any empirical data relevant to that assumption is a serious matter and should be remedied as soon as possible.

References

- Achenbach, T. M., & Edelbrock, C. S. (1978). The classification of child psychopathology: A review and analysis of empirical efforts. Psychological Bulletin, 85, 1275-1301.
- Adams, G. S. (1964). Measurement and evaluation in education, psychology, and guidance. New York: Holt, Rinehart & Winston.
- Alvino, J., McDonnell, R. C., & Richert, S. (1981). National survey of identification practices in gifted and talented education. Exceptional Children, 48, 124-132.
- American Psychological Association. (1974). The standards for psychological and educational tests. Washington, D.C.
- Anastasi, A. (1983). Psychological testing (5th edition). New York: Macmillan.
- Argulewicz, E. N., Elliott, S. N., & Hall, R. (1982). Comparison of behavioral ratings of Anglo-American and Mexican-American gifted children. Psychology in the Schools, 19, 469-472.
- Ashman, S. S., & Vukelich, C. (1983). The effect of different types of nomination forms on teachers' identification of gifted children. Psychology in the Schools, 20, 518-527.
- Baldwin, J. W. (1962). The relationship between teacher-judged giftedness, a group intelligence test and an individual test with possible gifted kindergarten pupils. The Gifted Child Quarterly, 6, 152-156.
- Baron, F., & Harrington, D. M. (1981). Creativity, intelligence, and personality. Annual Review of Psychology, 32, 439-476.
- Borko, H., & Cadwell, J. (1982). Individual differences in teachers' decision strategies: An investigation of classroom organization and management decisions. Journal of Educational Psychology, 74, 598-610.

- Borland, J. (1979). Teacher identification of the gifted: A new look. Journal for the Education of the Gifted, 2, 22-32.
- Borman, W. C. (1979). Format and training effects on rating accuracy and rater errors. Journal of Applied Psychology, 64, 410-421.
- Borthwick, B., Dow, I., Levesque, D., & Banks, R. (1980). The gifted and talented students in Canada: Results of a CEA survey. Toronto, Ontario: The Canadian Education Association.
- Burke, J. P., Haworth, C. E., & Ware, W. B. (1982). Scale for Rating Behavioral characteristics of Superior Students: An investigation of factor structure. Journal of Special Education, 16, 477-485.
- Callahan, C. M. (1982). Myth: There must be "winners" and "losers" in identification and programming. The Gifted Child Quarterly, 26, 17-19.
- Chambers, J. A., Barron, F., & Sprecher, J. W. (1980). Identifying gifted Mexican-American students. The Gifted Child Quarterly, 24, 123-128.
- Cook, T. D., & Campbell, D. T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Chicago: Rand McNally.
- Cornish, R. C. (1968). Parents', teachers', and pupils' perception of the gifted child's ability. The Gifted Child Quarterly, 12, 14-17.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. Psychological Bulletin, 52, 281-302.
- Cunningham, C. H., Thompson, B., & Alston, H. L. (1978). Use of S.O.I. abilities for prediction. The Gifted Child Quarterly, 22, 506-512.
- Davis, G. A., & Rimm, S. (1977). Characteristics of creatively gifted children. The Gifted Child Quarterly, 21, 546-551.
- Denton, C., & Postlethwaite, K. (1984). A study of the effectiveness of teacher-based identification of pupils with high ability in the secondary school. Gifted Education International, 2, 100-106.

- Dewing, K. (1970). The reliability and validity of selected tests of creative thinking in a sample of seventh-grade West Australian children. British Journal of Educational Psychology, 40, 35-42.
- Dirks, J., & Quarforth, J. (1981). Selecting children for gifted classes: Choosing for breadth vs choosing for depth. Psychology in the Schools, 18, 437-449.
- Edelbrock, C. (1979). Empirical classification of children's behavior disorders: Progress based on parent and teacher ratings. School Psychology Digest, 8, 355-369.
- Feldhusen, J. F., Asher, J. W., & Hoover, S. M. (1984). Problems in the identification of giftedness, talent, or ability. Gifted Child Quarterly, 28, 149-151.
- Feldhusen, J. F., Baska, L. K., & Womble, S. (1981). Using standard scores to synthesize data in identifying the gifted. Journal for the Education of the Gifted, 4, 177-186.
- Fox, L. H. (1981). Identification of the academically gifted. American Psychologist, 36, 1103-1111.
- Freeman, J. (1979). Gifted children: Their identification and development in a social context. Lancaster, England: MTP press.
- Gear, G. H. (1976). Accuracy of teacher judgment in identifying intellectually gifted children: A review of the literature. The Gifted Child Quarterly, 20, 478-490.
- Gear, G. H. (1978). Effects of training on teachers' accuracy in the identification of gifted children. The Gifted Child Quarterly, 22, 90-97.
- George, W. C. (1979). The Talent-Search concept: An identification strategy for the intellectually gifted. Journal of Special Education, 13, 221-237.

- Getzels, J. W., & Dillon, J. T. (1973). The nature of giftedness and the education of the gifted. In R. M. W. Travers (Ed.), Second handbook of research on teaching (pp. 689-731). Chicago: Rand McNally.
- Guilford, J. P. (1967). The nature of human intelligence. New York: McGraw-Hill.
- Hagen, E. (1980). Identification of the gifted. New York: Teachers College Press.
- Harrington, D. M., Block, J., & Block, J. H. (1983). Predicting creativity in preadolescence from divergent thinking in early childhood. Journal of Personality and Social Psychology, 45, 609-623.
- Harrington, R. G. (1982). Caution: Standardized testing may be hazardous to the educational programs of intellectually gifted children. Education, 103, 112-117.
- Hartsough, C. S., Elias, P., & Wheeler, P. (1983). Evaluation of a nonintellectual assessment procedure for the early screening of exceptionality. Journal of School Psychology, 21, 133-142.
- Hogan, R. (1980). The gifted adolescent. In J. Adelson (Ed.), Handbook of adolescent psychology. New York: Wiley.
- Hoge, R. D. (1983). Psychometric properties of teacher-judgment measures of pupil aptitudes, classroom behaviors, and achievement levels. Journal of Special Education, 17, 401-429.
- Hoge, R. D. (1984). The definition and measurement of teacher expectations: Problems and prospects. Canadian Journal of Education, 9, 213-228.
- Hoge, R. D., & Butcher, R. (1984). An analysis of teacher judgments of pupil achievement levels. Journal of Educational Psychology, 76, 777-781.

- Imada, A. S. (1982). Social interaction, observation, and stereotypes as determinants of differentiation in peer ratings. Organizational Behavior and Human Performance, 29, 397-415.
- Jacobs, J. A. (1971). Effectiveness of teacher and parent identifications of gifted children as a function of school level. Psychology in the Schools, 8, 140-142.
- Jenkins, R. A. (1979). A resource guide to preschool and primary programs for the gifted and talented. Mansfield Center, Conn.: Creative Learning Press.
- Karnes, F. A., & Collins, E. C. (1981). Assessment in gifted education. Springfield, Ill.: Charles Thomas.
- Kirk, W. D. (1966). A tentative screening procedure for selecting bright and slow children in kindergarten. Exceptional Children, 33, 235-240.
- Klausmeier, H. J., Harris, C. W., & Ethnathios, Z. (1962). Relationships between divergent thinking abilities and teacher ratings of high school students. Journal of Educational Psychology, 53, 72-75.
- Kornblau, B. (1982). The teachable pupil survey: A technique for assessing teachers' perceptions of pupil attributes. Psychology in the Schools, 19, 170-174.
- Love, K. G. (1981). Comparison of peer assessment methods: Reliability, validity, friendship bias, and user reaction. Journal of Applied Psychology, 66, 451-457.
- Lowenstein, L. F. (1982). Teachers' effectiveness in identifying gifted children. Gifted Education International, 1, 33-35.
- Mayfield, B. (1979). Teacher perception of creativity, intelligence, and achievement. The Gifted Child Quarterly, 23, 812-817.

- McIntyre, R. M., Smith, D. E., & Hassett, C. E. (1984). Accuracy of performance ratings as affected by rater training and perceived purpose of rating. Journal of Applied Psychology, 69, 147-156.
- Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. Psychological Bulletin, 52, 194-216.
- Messick, S. (1981). Constructs and their vicissitudes in educational and psychological measurement. Psychological Bulletin, 89, 575-588.
- Murphy, D., Jenkins-Friedman, R., & Tollefson, N. (1984). A new criterion for the 'ideal' child? The Gifted Child Quarterly, 28, 31-36.
- Pegnato, C. W., & Birch, J. W. (1959). Locating gifted children in junior high school - a comparison of methods. Exceptional Children, 25, 300-304.
- Pledgie, T. K. (1982). Giftedness among handicapped children: Identification and programming development. Journal of Special Education, 16, 221-227.
- Quay, H. C. (1979). Classification. In H. C. Quay, & J. S. Werry (Eds.), Psychopathological disorders of childhood, 2nd ed. (pp. 1-42). New York: Wiley.
- Renzulli, J. S. (1978). What makes giftedness? Reexamining a definition. Phi Delta Kappan, 60, 180-184.
- Renzulli, J. S. (1984, April). What makes giftedness? A reexamination of major constructs and approaches to identification. Paper presented at meeting of American Educational Research Association, New Orleans, La.
- Renzulli, J. S., & Hartman, R. K. (1971). Scale for Rating Behavioral Characteristics of Superior Students. Exceptional Children, 38, 243-248.
- Renzulli, J. S., Hartman, R. K., & Callahan, C. M. (1971). Teacher identification of superior students. Exceptional Children, 38, 211-214.

- Renzulli, J. S., Reis, S. M., & Smith, L. H. (1981). The revolving door identification model. Mansfield, Conn.: Creative Learning Press.
- Robinson, H. B., Roedell, W. C., & Jackson, N. E. (1979). Early identification and intervention. In A. H. Passow (Ed.), The gifted and the talented: Their education and development (pp. 138-154). Chicago, Ill.: University of Chicago Press.
- Rosenfield, S. (1983). Assessment of the gifted child. In T. R. Kratochwill (Ed.), Advances in school psychology, vol. 3 (pp. 141-174). Hillsdale, N.J.: L. Erlbaum.
- Rubensz, R. (1979). Identification and evaluation procedures for gifted and talented programs. The Gifted Child Quarterly, 23, 304-313.
- Rubin, K. H., & Clark, M. L. (1983). Preschool teacher's ratings of behavioral problems: Observational, sociometric, and social-cognitive correlates. Journal of Abnormal Child Psychology, 11, 273-286.
- Rust, J. O., & Lose, B. D. (1980). Screening for giftedness with the Slossen and the Scale for Rating Behavioral Characteristics of Superior Students. Psychology in the Schools, 17, 446-451.
- Sattler, J. M. (1982). Assessment of children's intelligence and special abilities. Boston, Mass.: Allyn and Bacon, Inc.
- Schlichter, C. L. (1981). The multiple talent approach in mainstream and gifted programs. Exceptional Children, 48, 144-150.
- Shavelson, R. J., Cadwell, J., & Izu, T. (1977). Teachers' sensitivity to the reliability of information in making pedagogical decisions. American Educational Research Journal, 14, 83-97.
- Stanley, J. C. (1976). Test better finder of great math talent than teachers are. American Psychologist, 31, 313-314.

- Stanley, J. C., Keating, D. P., & Fox, L. H. (1974). Mathematical talent: Discovery, description, and development. Baltimore, Md.: Johns Hopkins University Press.
- Sternberg, R. J. (in press). Toward a triarchic theory of human intelligence. Behavioral and Brain Sciences.
- Swenson, E. V. (1978). Teacher-assessment of creative behavior in disadvantaged children. The Gifted Child Quarterly, 22, 338-343.
- Tannenbaum, A. (1979). Pre-Sputnik to post-Watergate concern about the gifted. In A. H. Passow (Ed.), The gifted and the talented: Their education and development. Chicago, Ill.: Univ. of Chicago Press.
- Terman, L. M. (1925). Mental and physical traits of a thousand gifted children: Genetic studies of genius (Vol. 1). Stanford, Calif.: Stanford Univ. Press.
- Torrance, E. P. (1965). Rewarding creative behavior. Englewood Cliffs, N.J.: Prentice-Hall.
- Treffinger, D. J. (1984, April). Gifted education: Time for a reappraisal? Paper presented at meeting of American Educational Research Association, New Orleans, La.
- Treffinger, D. J., Pyrry, M. C., Hawk, M. M., & Houseman, E. D. (1979). Education of the gifted and talented: Implications for school psychology. In G. D. Phye and D. J. Reschly (Eds.), School psychology: Perspectives and issues (pp. 191-214). New York: Academic Press.
- Tuttle, F. B., & Becker, L. A. (1980). Characteristics and identification of gifted and talented students. Washington, D.C.: National Education Association.

United States Department of Education. (1983). Meeting the challenge:

Recent efforts to improve education across the nation. Washington, D.C.:

U.S. Government Printing Office.

Whitmore, J. R. (1980). Giftedness, conflict, and underachievement. Boston,

Mass.: Allyn and Bacon.

Wilson, C. D. (1963). Using test results and teacher evaluation in

identifying gifted pupils. Personnel and Guidance Journal, 41, 720-721.

Yarborough, B. H., & Johnson, R. A. (1983). Identifying the gifted: A

theory-practice gap. The Gifted Child Quarterly, 27, 135-138.

Zedek, S., & Cascio, W. F. (1982). Performance appraisal decisions as a

function of rater training and purpose of the appraisal. Journal of

Applied Psychology, 67, 752-758.

Author Notes

Support for this project was provided by a grant from the Pickering Fund. We are grateful to B. Wickert, D.A. Andrews, J.S. Renzulli, and D.J. Treffinger for their comments on an earlier version of this manuscript. Versions of this paper were presented at the 1985 Annual Conference of the Canadian Psychological Association and at the Sixth World Conference on Gifted and Talented Children.

Footnotes

¹Further developments of the "gifted" construct will, of course, come from a number of directions, including better models of intellectual functioning (e.g., Hogan, 1980; Sternberg, in press).

²This construct validity research should also attempt to treat the teacher as a unit of analysis in an effort to determine whether or not there are individual differences with respect to judgmental accuracy (see, for example, Borko & Cadwell, 1982; Denton & Postlethwaite, 1984; Hoge & Butcher, 1984).

Table 1

Summary of Validity Data

Investigation	Grade Level	N	Judgment Measure	Criterion Measure	Design/Analysis	Results
Ashman & Vukelich (1983)	K-5	183	a. Ratings	IQ Test	Effectiveness ^a	20% - 81%
					Efficiency	54% - 71%
			b. Nominations	IQ Test	Effectiveness	33%
					Efficiency	78%
Baldwin (1962)	K	140	Nominations	IQ Test	Efficiency	26% - 38%
Borland (1979)	3,4,66	195	Ratings	IQ Test	Correlation ^b	r = .22, .32
Chambers, Baron, & Sprecher (1980)	3-6	298	Ratings	IQ Test	t-test	t = 3.06 - 4.05
				Creativity	t-test	t = 2.16 - 4.37
Cornish (1968)	6	86	Nominations	IQ Test	Effectiveness	31%
					Efficiency	42%
Cunningham, Thompson, & Alston (1978)	4-6	138	Ratings	Creativity	Regression ²	R = .20 - .31
Davis & Rimm (1977)	1-6	365	Ratings	Creativity	Correlation	r = .30
Denton & Postlethwaite (1984)	7-8	NA	Nomination	Achievement	Effectiveness	69% - 86%
Dewing (1970)	7	394	Nomination	Creativity	Chi-Square	$\chi^2 = 2.91 - 81.48$
Gear (1978)	3-6	NA	Nomination	Expert	Effectiveness	40% - 86%
					Efficiency	19% - 24%

Investigation	Grade Level	N	Judgment Measure	Criterion Measure	Design/Analysis	Results
Harrington, Block, & Block (1983)	6	80	Ratings	Creativity	Correlation	$r = .45$
Hartsough, Elias, & Wheeler (1983)	K-2	536	Nomination	IQ Test	Effectiveness	0%
Jacobs (1971)	K	654	Nomination	IQ Test	Effectiveness	10%
					Efficiency	4%
Kirk (1966)	K	112	Ratings	IQ Test	Correlation ^a	$r = .41 - r = .73$
Klausmeier, Harris & Echnathios (1962)	10&11	191	Ratings	Creativity	Correlation	$r = .05 - r = .70$
Lowestein (1982)	1-12	163	Nomination	IQ Test	Effectiveness	70%
					Efficiency	69%
Mayfield (1979)	3	573	Ratings	Creativity	Correlation	c
Pegnato & Birth (1959)	7-12	781	Nomination	IQ Test	Effectiveness	45%
					Efficiency	27%
Renzulli, Hartman, & Callahan (1971)	4-6	72	Ratings	IQ Test	Correlation	$r = .36$ & $r = .61$
				Achievement	Correlation	$r = .41$ to $r = .60$
				Creativity	Correlation	$r = .24$ to $r = .48$
Rust & Lose (1980)	1-7	438	Ratings	IQ Test	Correlation	$r = .01 - r = .20$

Investigation	Grade Level	N	Judgment Measure	Criterion Measure	Design/Analysis	Results
Swenson (1978)	4-6	90	Ratings	Achievement	Correlation	$r = .39$
				Creativity	Correlation	$r = .08$
Wilson (1963)	2-12	205	Nomination	IQ Test	Effectiveness	45%

- a. Correlational analyses are also reported. b. Effectiveness/efficiency analyses are also reported.
 c. Correlations are reported as significant but coefficients are not provided.

Abstract

There is considerable emphasis today on the provision of special educational treatment for academically gifted pupils. A variety of selection tools are used in the identification of such pupils, including intelligence tests, achievement tests, creativity measures, and teacher-judgment measures. It is the latter type of measure which forms the focus of this review, and the purpose is to assess the psychometric properties of these teacher-judgment measures in terms of the available empirical data. The major conclusion from the review is that there is little basis for the negative assessments so often associated with these measures. The paper also includes a set of recommendations regarding the use of the measures in the identification of gifted pupils and a set of recommendations regarding future research on the measures.